Lecture 5:
Online Experiments

Raymond Duch
Official Fellow Nuffield College
Director CESS Nuffield/Santiago

CENTRE FOR EXPERIMENTAL SOCIAL SCIENCES

Road Map to Presentation

- Review characteristics of online experiments
- Why online experiments?
- Subjects pools
- Technologies
- Randomisation
- Example
- Conjoint Experiments
- Mode effects

4 Reasons for leaving the lab

- A more diverse subject pool
- Power
- Co-variates
- Cost

CENTRE FOR EXPERIMENTAL SOCIAL SCIENCES

A more diverse subject pool?

- Internal validity pertains to the question of whether causal estimates approximately reflect the effects of the experimental manipulation among the participants in the original setting. Were subjects treated?
    - So maybe no.
- External validity is an assessment of whether the casual estimates deduced from experimental research would persist in other settings and with other samples.
    - So maybe yes.

Belot et al 2015

- Students are more likely to behave as homo-economicus agents than non-students in games involving other-regarding preferences (Dictator Game, Trust Game and Public Good Game).

- Students are highly consistent in their other-regarding preferences while non-student subjects are muchmore inconsistent across other-regarding games.

Students vs Non-students

Aggregate behavior in the five games.

Game	Decision	EB-NS	EB-S	*p*-Value	D-NS	D-S	*p*-Value
Dictator	0–10	17	57	<0.001	3.5	1.6	<0.001
Trust—sender	0–10	18	65	<0.001	8.2	3.5	<0.001
Trust—receiver	0–15	13	44	=0.005	13.1	8.4	=0.005
Public Good r1	20 tokens MRC = 0.4	9	24	=0.026	11.1	9.9	=0.323
Public Good r10	20 tokens MRC = 0.4	29	52	=0.008	5.4	4.8	=0.151
Guessing	20 if closest to 2/3 of average (0–100)	32	56	=0.008	45.9	38.3	=0.051
Auction	Private value: 4	29	27	=0.908	4.8	3.7	=0.205
Auction	Private value: 6	46	31	=0.420	5.1	4.9	=0.730
Auction	Private value: 8	13	62	=0.005	5.6	7.7	=0.008
Auction	Private value: 10	18	36	=0.175	7.3	7.7	=0.311

Note: *p*-values obtained from two-sample test of proportions and two-sample ranksum test. EB-(N)S: % equilibrium behavior of non(students). D-(N)S: actual decision of (non)students.

Students vs Non-students

Demographics and behavior—probability of equilibrium decision probit estimates—estimated student dummy coefficient.

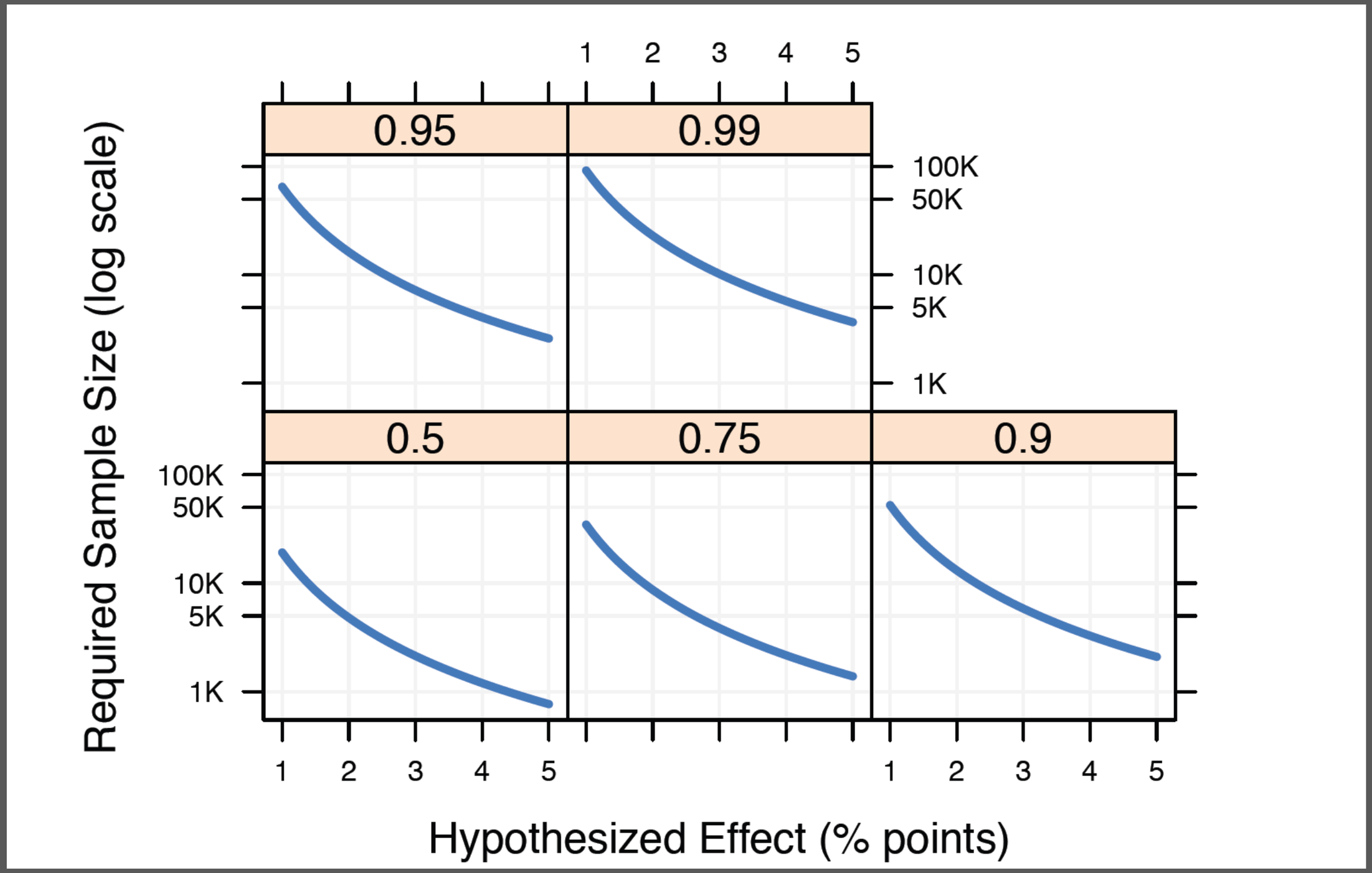
	(1)	(2)	(3)	(4)
(1) Dictator game	.40***	.41***	.31*	.31*
	(.11)	(.12)	(.16)	(.16)
(2) Trust game—sender	.46***	.50***	.55***	.55***
	(.11)	(.12)	(.14)	(.14)
(3) Trust game—receiver	.31***	.22*	.38**	.38***
	(.11)	(.12)	(.14)	(.14)
(4) Public good game	.15**	.10	.10	.10
(round 1)	(.06)	(.07)	(.08)	(.08)
(5) Guessing game	.23***	.19**	.06	.06
	(.09)	(.09)	(.11)	(.12)
(6) Auction	.12	.11	.21*	.21*
	(.08)	(.09)	(.10)	(.10)
Control for cognitive ability	No	Yes	Yes	Yes
Control for age	No	No	Yes	Yes
Control for gender	No	No	No	Yes

Note: Each cell presents the estimates of the student dummy corresponding to a separate regression. ***, **, and * denote significance levels at 1, 5, and 10%.

Remember power calc

- Type II Error: False Negative
- Beta is probability of Type II Error
    - if the null is not true, how often can we reject the null successfully?
- 1-Beta is power of a test: probability that the test rejects Ho
- as N increases the power of the test increases

Power Curves: different power levels



Power Curves: different power levels

Table : The Power of Numbers: Lab versus Internet

	Lab	Online
Subjects	200	1000
Control Group Outcome	60	60
Treatment Group Outcome	65	65
Standard Deviation	20	20
Significance	0.05	0.05
1-Beta	0.42	0.98

Covariates & heterogeneity

- You have a conjecture regarding treatment effects
- Co-variates are unimportant with random assignment
- Do you have a conjecture re: heterogeneity?
- Can “covariates” be incorporated into design?
- Are there relevant characteristics that subjects bring to the experiment?

Your Subject Sample

- Probability Sample
  - Knowledge Networks
- Representative Sample
  - Respondi
- Diverse Sample
  - M-Turk
  - CESS UK/CESS Latin America/CESS India/CESS China
- Lab Convenience Sample
  - CESS Lab Subject Pool

An Internet Probability Sample

- **Probability based panel sample that is representative of the U.S. population.**
- **Panel sample is selected using high quality RDD sampling methods – comparable to those used by the U.S. government's RDD surveys (CDC's National Immunization Survey, for example).**
- **Other Internet research that covers only individuals with Internet access who volunteer for research,**
- **Knowledge Networks surveys are based on a sampling frame that includes both listed and unlisted phone numbers, and is not limited to current Web users or computer owners. Panelists are selected by chance to join the panel;**
- **unselected volunteers are not able to join the KN panel.**

Improving Representativeness

- Polimetrix/Doug Rivers — Sample selection by matching
- Enumeration of target population — commercial/government exhaustive lists
- Draw a stratified random sample from the target population using exhaustive lists
- Choose matched sample:
    - exact matching
    - propensity score matching
    - proximity matching

E.g., proximity matching

- Minimize distance between each member of the target sample and a member of panel

- Weighted sum of the individual distance functions on each attribute

- d(x,y) = |x-y|

Efficient Randomisation and Causal Heterogeneity

Horiuchi et al 2007

- Experimental subjects choose not to comply/not to respond
- Ignoring this selection leads to invalid causal inferences
- Therefore design randomised experiments that allow for statistical adjustments of noncompliance and nonresponse
- We now examine the Horiuchi et al approach

Efficient Randomisation

Information and Voter Turnout in Japan

- Experiment manipulates quality and quantity of info voters receive

- 6000 internet for pre-screening questions – 2,748 answer

- 2000 randomly selected for internet experiment

- random assignment of these to treatments

  - one-party treatment sent to either LDP or DPJ website
  - two-party treatment sent to websites of both parties – order randomised
  - control group not sent to any website

	Randomized Blocks						
	I	II	III	IV	V	VI	
	Planning to Vote		Not Planning to Vote		Undecided		
	Male	Female	Male	Female	Male	Female	Total
One-party treatment group							
DPJ website	194	151	24	33	36	62	500
LDP website	194	151	24	33	36	62	500
Two-party treatment group							
DPJ/LDP websites	117	91	15	20	20	37	300
LDP/DPJ websites	117	91	15	20	20	37	300
Control group							
no website	156	121	19	26	29	49	400
Block size	778	605	97	132	141	247	2000

		Treatment Assignment	
		$Z_i = 1$	$Z_i = 0$
Actual Treatment Received	$T_i = 1$	**Complier** $Y_i(1)$ is observable; $C_i = 1$	
	$T_i = 0$	**Noncomplier** $Y_i(1)$ is observable; $C_i = 0$	**Complier** or **Noncomplier** $Y_i(0)$ is observable; $C_i =??$

Intention to Treat Effect (ITT)

- the ITT effect is the causal effect of being asked to visit the party website
- its NOT the effect of the actual treatment received
- political parties cannot force every voter to visit website – so this is of interest

$$\text{ITT} = \frac{1}{N} \sum_{i=1}^{N} [Y_i(1) - Y_i(0)]$$

Complier Average Causal Effect (CACE)

- focus on causal effect of the treatment for compliers in upper left and lower right cells
- of interest is sample complier average causal effect or CACE

$$\frac{\sum_{i=1}^{N} C_i[Y_i(1)-Y_i(0)]}{\sum_{i=1}^{N} C_i}$$

Modeling Compliance and Non-response

- First model the conditional probability of being a complier

$$\Pr(C_i = 1 | X_i, \xi) = \Phi(X_i^\top \xi)$$

- Second model voter turnout given the compliance status, C_i, treatment assignment Z_i, and the observed covariates X_i

$$\Pr(Y_i = 1 | C_i, Z_i, X_i, \alpha, \beta, \gamma) = \Phi[\alpha C_i Z_i + \beta C_i (1 - Z_i + X_i^\top \gamma]$$

Sequential Randomised Experiments

Types of Sequential Randomised Experiments

- Non-adaptive – assigment probabilities fixed
- Treatment-adaptive – change based on number subjects in treatment
- Covariate-adaptive – change based on covariate profiles of new and previous subjects
- Responsive-adaptive – change as function of previous units outcomes

Biased coin method with discrete covariates

- t, letting $t \in (1, 2)$ corresponds to treatment control and treatment conditions

- a unit enters the experiment but prior to randomisation

- J discrete covariates are measured indexed by j

- the jth covariate has l_j levels, indexed by $i \in (1, ..., l_j)$

- when the current subject arrives, all previous subjects' covariate values and treatment assignments are known

Randomisation protocol

- $$S_p = sgn(n_{ij1} - n_{ij2})$$
- if $S_p > 0$ assign the current subject to treatment with some probability $\pi > \frac{1}{2}$
- some consensus that $\frac{2}{3}$ or $\frac{3}{4}$
  - if $S_p = 0$ let $\pi = \frac{1}{2}$
  - if $S_p < 0$ let $\pi < \frac{1}{2}$

Table 2 Consistent marginal and joint distributions of two binary covariates

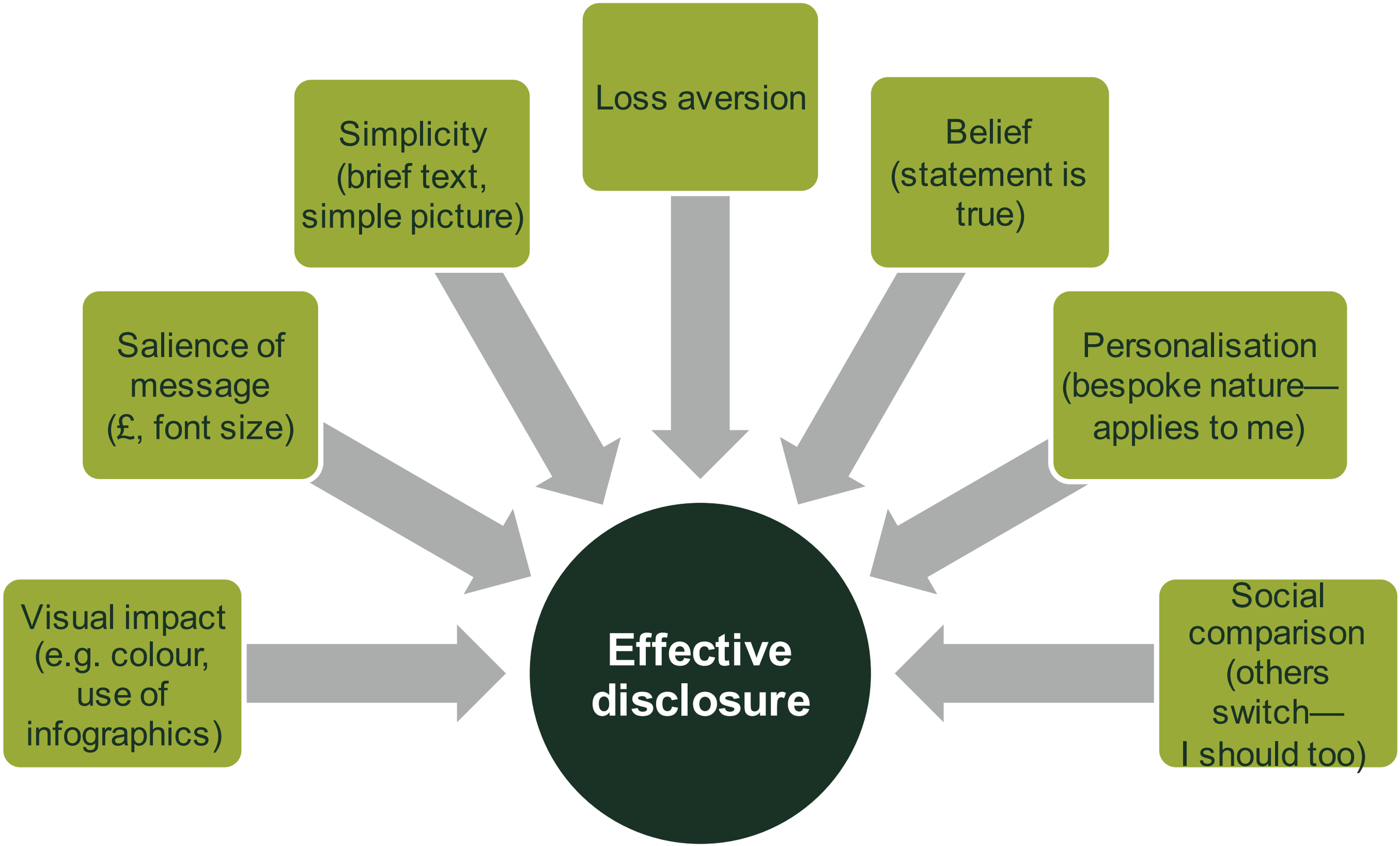
	Sex		*Party*	
	M	*F*	*Rep*	*Dem*
Control	3	3	3	3
Treatment	3	3	3	3

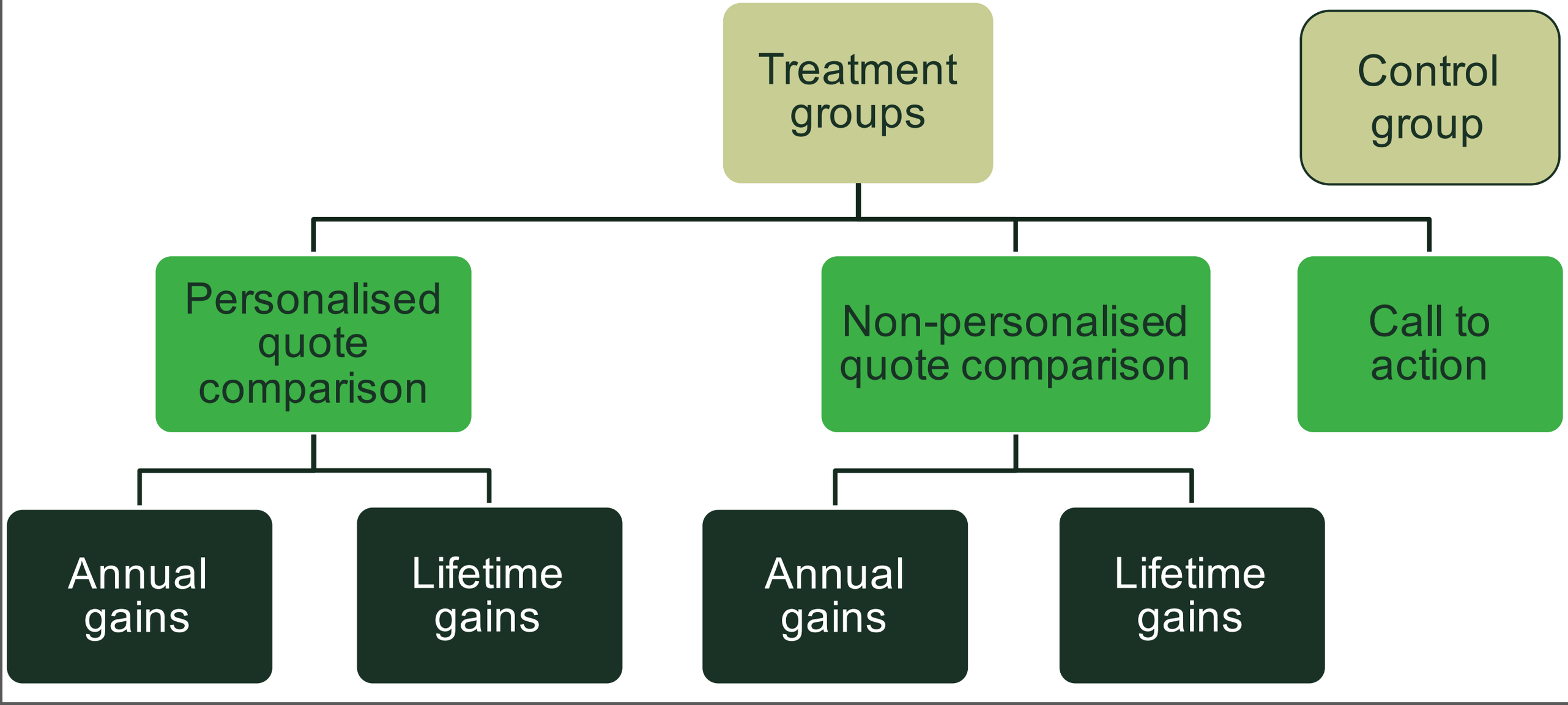
	Rep M	*Dem M*	*Rep F*	*Dem F*
Control	3	0	0	3
Treatment	0	3	3	0

Note. The upper panel reflects the minimization approach and suggests balance between treatment conditions. However, the lower panel makes clear that results will represent isolated parts of the covariate support.

Internet Experiments @ CESS

- Online subject pool
  - Third party subject pools
  - UK CESS online subject pool
    - Santiago CESS online subject pool
- Programming
  - Qualtrics
  - Customized (FCA Example)


Loss aversion
Simplicity (brief text, simple picture)
Belief (statement is true)
Salience of message (£, font size)
Personalisation (bespoke nature—applies to me)
Visual impact (e.g. colour, use of infographics)
Effective disclosure
Social comparison (others switch—I should too)


Treatment groups
Control group
Personalised quote comparison
Non-personalised quote comparison
Call to action
Annual gains
Lifetime gains
Annual gains
Lifetime gains

Time elapsed **0:22**

Your pension provider

Your pension provider

Annuity features

- Pension pot **£24,597**
- Paid **quarterly**
- Paid in **advance**
- **Single** annuity
- **5 years** guarantee period
- Increase by **inflation**

These cannot be changed

Our quote for this product

The annuity product offered by us would provide you with an annual income of:

£1,379

It's not too late to shop around for quotes from other providers.

Purchase our product

We are required by the Financial Conduct Authority to inform you that you can shop around if you want to. If you want to see what other options are available from other providers please click here and you will be taken to a secure comparison site. Other providers will not know all necessary information about you or your circumstances. In order to shop around, you will need to provide **personal information**, including that relating to your health and lifestyle.

Control

Time elapsed **0:34**

Your pension provider

Your pension provider

Annuity features

- Pension pot **£24,597**
- Paid **quarterly**
- Paid in **advance**
- **Single** annuity
- **5 years** guarantee period
- Increase by **inflation**

These cannot be changed

Our quote for this product

The annuity product offered by us would provide you with an annual income of:

£1,379

Can you get a better income from your annuity?

Based on your key information, there are quotes available from other providers offering higher rates. If you select our product you would be losing out on **£46** a year.

Purchase our product

We are required by the Financial Conduct Authority to inform you that you can shop around if you want to. If you want to see what other options are available from other providers please click here and you will be taken to a secure comparison site. Other providers will not know all necessary information about you or your circumstances. In order to shop around, you will need to provide **personal information**, including that relating to your health and lifestyle.

Personalised Quote Comparison

Centre for Experimental Social Sciences

Time elapsed **0:27**

Your pension provider

Your pension provider

Annuity features

- Pension pot **£24,597**
- Paid **quarterly**
- Paid in **advance**
- **Single** annuity
- **5 years** guarantee period
- Increase by **inflation**

These cannot be changed

Our quote for this product

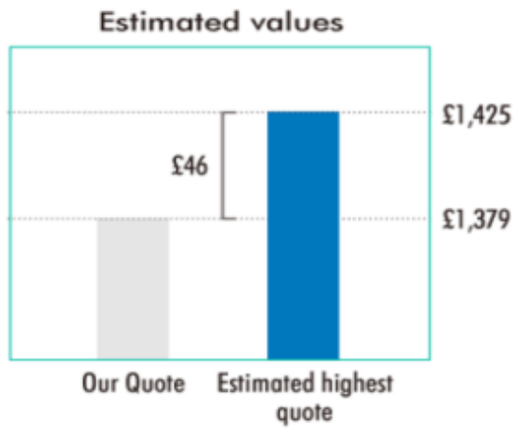
The annuity product offered by us would provide you with an annual income of:

£1,379

Can you get a better income from your annuity?

Based on your key information, we have estimated the highest annuity income you might be offered by other providers. Based on this estimate, if you select our product, you might lose out on around **£50** a year.

This estimate does not use real-time quotes from other annuity providers. As a result, the estimate may be higher or lower than the annuity quotes you would actually be offered, were you to shop around.



Purchase our product

We are required by the Financial Conduct Authority to inform you that you can shop around if you want to. If you want to see what other options are available from other providers please click here and you will be taken to a secure comparison site. Other providers will not know all necessary information about you or your circumstances. In order to shop around, you will need to provide **personal information**, including that relating to your health and lifestyle.

Non-personalised Quote Comparison

Time elapsed **0:11**

Your pension provider

Your pension provider

Annuity features

- Pension pot **£24,597**
- Paid **quarterly**
- Paid in **advance**
- **Single** annuity
- **5 years** guarantee period
- Increase by **inflation**

These cannot be changed

Our quote for this product

The annuity product offered by us would provide you with an annual income of:

£1,379

Can you get a better income from your annuity?

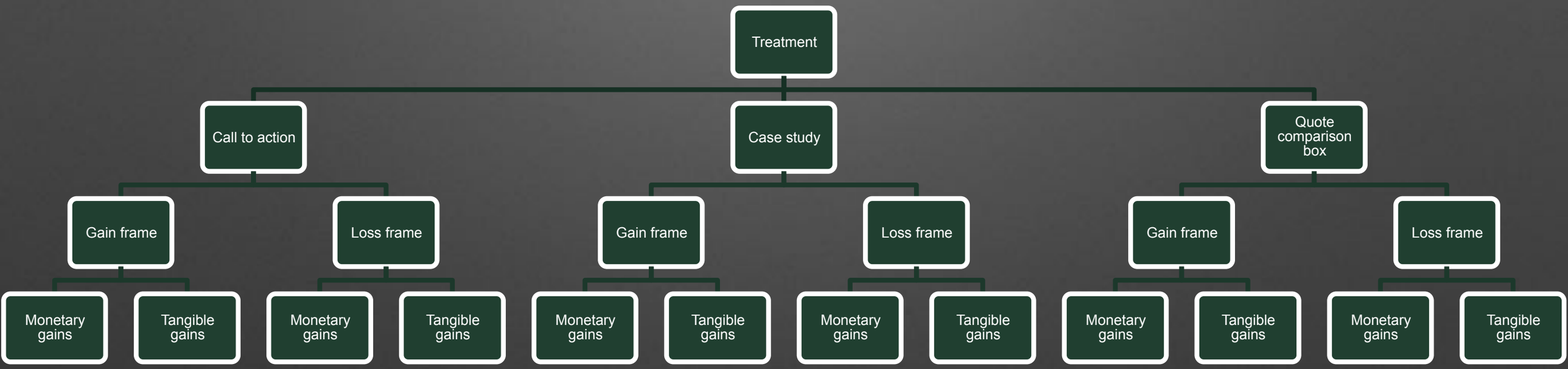
80% of people lose out by purchasing from their own pension provider according to a 2014 survey.

Purchase our product

We are required by the Financial Conduct Authority to inform you that you can shop around if you want to. If you want to see what other options are available from other providers please click here and you will be taken to a secure comparison site. Other providers will not know all necessary information about you or your circumstances. In order to shop around, you will need to provide **personal information**, including that relating to your health and lifestyle.

Call to Action

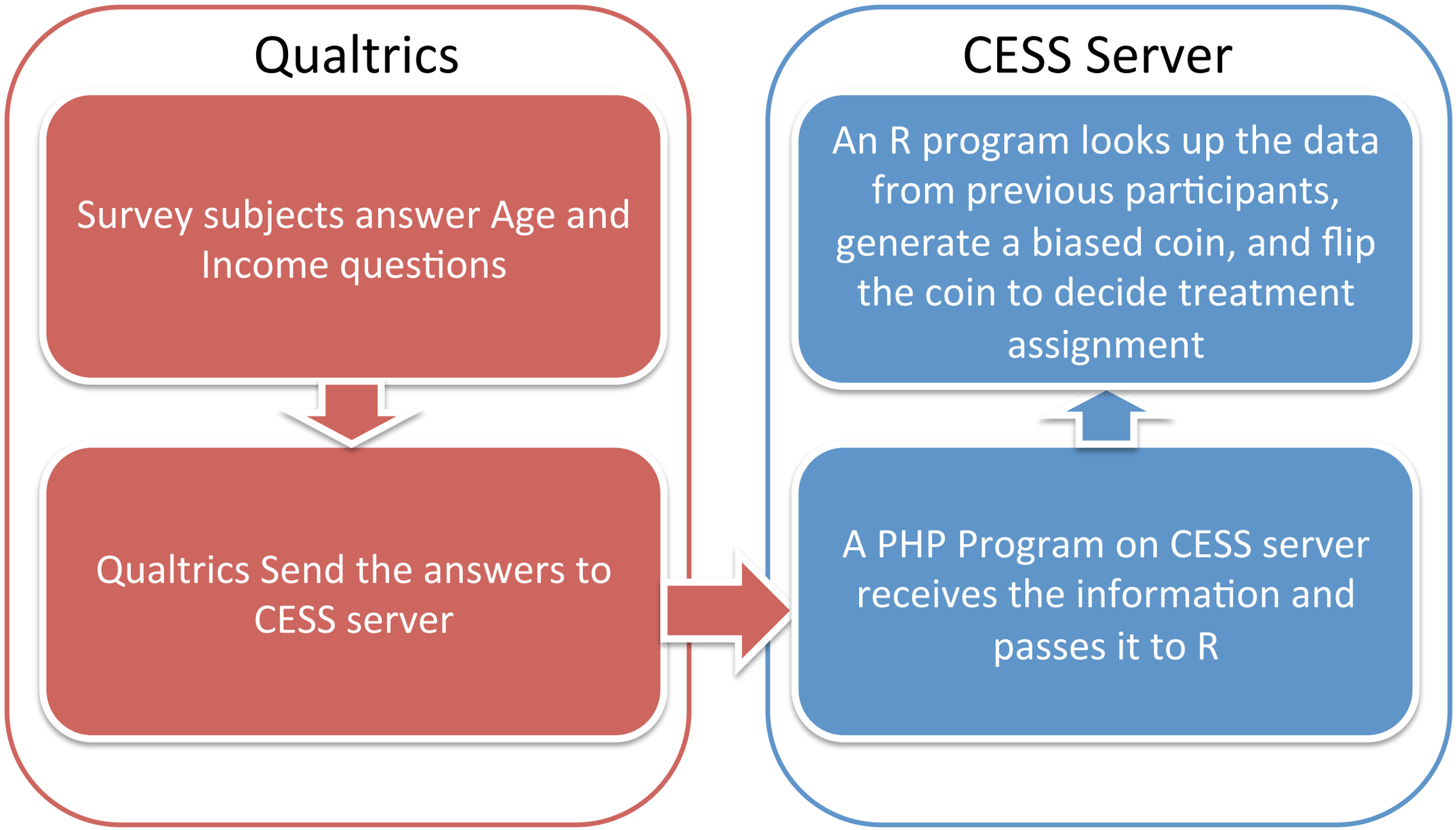
FCA: Assignment to Treatments



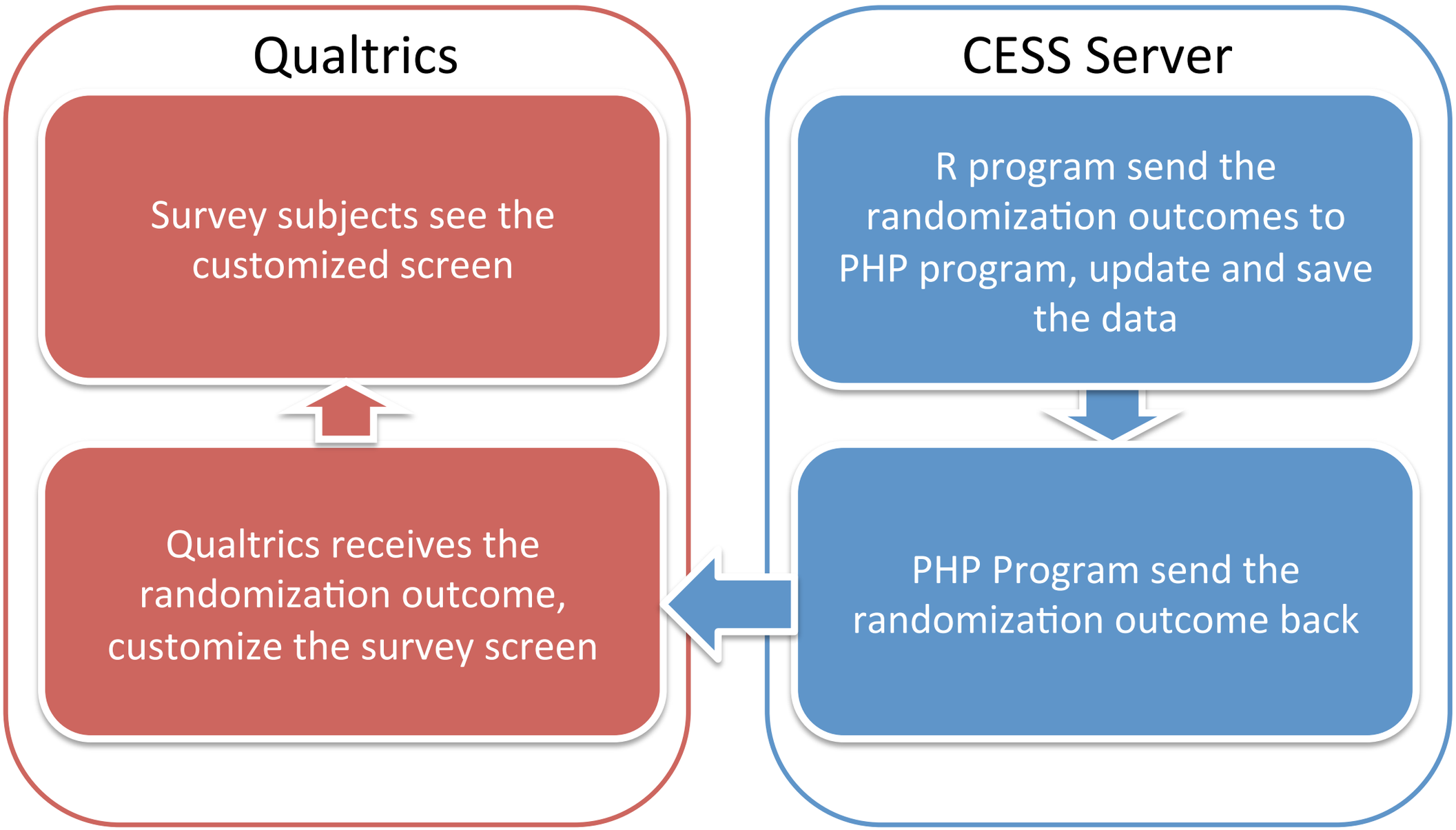
How to Assign to Treatment?

- Block random assignment (sequential online)
- Subjects are partitioned into blocks
- Complete random assignment within each block
- Example
  - Split subjects by gender
  - Select 5 subjects from the male group for the treatment
  - Select 5 subjects from the female group for the treatment

Step 1


Qualtrics
Survey subjects answer Age and Income questions
Qualtrics Send the answers to CESS server
CESS Server
An R program looks up the data from previous participants, generate a biased coin, and flip the coin to decide treatment assignment
A PHP Program on CESS server receives the information and passes it to R

Step 2


Qualtrics
Survey subjects see the customized screen
Qualtrics receives the randomization outcome, customize the survey screen
CESS Server
R program send the randomization outcomes to PHP program, update and save the data
PHP Program send the randomization outcome back

Conjoint Experiments

- Typical survey experiments test uni-dimensional causal effect
    - Treatment versus Control
    - Typical vignette experiments: immigration, candidate, racial cue
    - But what components of manipulation produce observed effect?
        - Why does immigration status matter?

Hainmueller et al

- Political Analysis 2013 Causal Inference in Conjoint Analysis
- Understanding Multidimensional Choices via State Preference Experiments
- The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes towards Immigrants
- Validating Vignette and Conjoint Survey Experiments Against Real-World Behaviour

Example: Candidates

- Respondents chose between/rank 2 candidates
    - J=2 (choices)
    - K=6 choices/evaluations
- Each candidate has a profile
    - Each profile has set of L discretely valued attributes, or a treatment composed of L components.
    - We use D_l to denote the total number of levels for attribute l
    - **L**=8 (candidate attributes), D(1)…**D(**6) (total number of levels for candidate's age, education, etc.), and D(7) …D(8)=2 (for military service and gender)

CENTRE FOR EXPERIMENTAL SOCIAL SCIENCES

Please read the descriptions of the potential immigrants carefully. Then, please indicate which of the two immigrants you would personally prefer to see admitted to the United States.

	Immigrant 1	Immigrant 2
Prior Trips to the U.S.	Entered the U.S. once before on a tourist visa	Entered the U.S. once before on a tourist visa
Reason for Application	Reunite with family members already in U.S.	Reunite with family members already in U.S.
Country of Origin	Mexico	Iraq
Language Skills	During admission interview, this applicant spoke fluent English	During admission interview, this applicant spoke fluent English
Profession	Child care provider	Teacher
Job Experience	One to two years of job training and experience	Three to five years of job training and experience
Employment Plans	Does not have a contract with a U.S. employer but has done job interviews	Will look for work after arriving in the U.S.
Education Level	Equivalent to completing two years of college in the U.S.	Equivalent to completing a college degree in the U.S.
Gender	Female	Male

	Immigrant 1	Immigrant 2
If you had to choose between them, which of these two immigrants should be given priority to come to the United States to live?	○	○

On a scale from 1 to 7, where 1 indicates that the United States should absolutely not admit the immigrant and 7 indicates that the United States should definitely admit the immigrant, how would you rate Immigrant 1?

Absolutely Not Admit 1	2	3	4	5	6	Definitely Admit 7
○	○	○	○	○	○	○

Using the same scale, how would you rate Immigrant 2?

Absolutely Not Admit 1	2	3	4	5	6	Definitely Admit 7
○	○	○	○	○	○	○

Fig. 1 Experimental design: Immigration conjoint. This figure illustrates the experimental design for the conjoint analysis that examines immigrant admission to the United States.

Respondent's JK Profiles

- $Y_{ijk}(\bar{\mathbf{T}}_\mathbf{i})$

$$\sum_{j=1}^{J} Y_{ijk}(\bar{\mathbf{t}}) = 1$$

Stability Assumption

- potential outcomes always take on the same value as long as all the profiles in the same choice task have identical sets of attributes.

$$Y_{ijk}(\overline{\mathbf{T}}_{\mathbf{i}}) = Y_{ijk^t}(\overline{\mathbf{T}}_{\mathbf{i}})$$

if $\overline{\mathbf{T}}_{\mathbf{ik}} = \overline{\mathbf{T}}_{\mathbf{ik}^t}$

for any j, k, k^t

No Profile Order Effect

$$Y_{ij}(\mathbf{T_{ik}}) = Y_{ij'}(\mathbf{T_{ik}}')$$

if $T_{ijk} = T'_{ij'k}$

and $T_{ij'k} = T'_{ijk}$

for any i, j, j', k

Randomisation of Profiles

- $Y_i(\mathbf{t}) \perp\!\!\!\perp T_{ijkl},$ for any i, j, k, l
- pairwise independence between all elements of $Y_i(\mathbf{t})$ and T_{ijkl}
- and $0 < p(\mathbf{t}) = p(\mathbf{T_{ik}} = \mathbf{t}) < 1)$

Basic Profile Effects

$$\pi(\mathbf{t_1}, \mathbf{t_0}) = Y_i(\mathbf{t_1}) - Y_i(\mathbf{t_0})$$

Profiles	Candidate	Service	Income	Education
t_0	1	military	rich	college
	2	no service	poor	college
t_1	1	military	rich	college
	2	military	poor	college

Estimate Profile Effects

- Unit-level causal effects are difficult to identify
- Involve counterfactuals and hence fundamental problem of causal inference
- Average Treatment Effects (ATE)?
- If there are a large number of attributes with multiple levels the number of observations in each conditioning set will be virtually zero rendering estimation difficult if not impossible

Avg Marginal Component Effect

$$\hat{\bar{\pi}}_l(t_1, t_0, p(\boldsymbol{t})) = \sum_{(t,\mathbf{t})\in\tilde{\mathcal{T}}} \Big\{ \mathbb{E}\left[Y_{ijk} | T_{ijkl} = t_1, T_{ijk[-l]} = t, \boldsymbol{T}_{i[-j]k} = \boldsymbol{t}\right]$$

$$- \mathbb{E}\left[Y_{ijk} | T_{ijkl} = t_0, T_{ijk[-l]} = t, \boldsymbol{T}_{i[-j]k} = \boldsymbol{t}\right] \Big\}$$

$$\times p\Big(T_{ijk[-l]} = t, \boldsymbol{T}_{i[-j]k} = \boldsymbol{t} | (T_{ijk[-l]}, \boldsymbol{T}_{i[-j]k}) \in \tilde{\mathcal{T}}\Big)$$

- the marginal effect of attribute l averaged over the joint distribution of the remaining attributes

Estimating AMCE

- for any attribute of interest T_{ijkl} the subclassification estimate of the AMCE can be computed simply by dividing the sample into the strata defined by T_{ijk}

- typically the attributes on which the assignment of the attribute of interest is restricted

- calculate the difference in the average observed choice outcomes between the treatment ($T_{ijkl} = 1$) and control ($T_{ijkl} = 0$) groups within each stratum

- take the weighted average of these differences in means, using the known distribution of the strata as the weights

Regression Estimation

- the linear regression estimator is fully nonparametric, even though the estimation is conducted by a routine typically used for a parametric linear regression model

- regress the outcome variable on the L sets of dummy variables

- interaction terms for the attributes that are involved in any of the randomization restrictions used in the study

- take the weighted average of the appropriate coefficients

Variance Estimation

- observed choice outcomes within choice tasks strongly negatively correlated
- both potential choice and rating outcomes within respondents are likely to be positively correlated because of unobserved respondent characteristics influencing their preferences
- point estimates of the AMCE can be coupled with standard errors corrected for within respondent clustering
- obtain cluster-robust standard errors for the estimated regression coefficients by using the cluster option in Stata
- block bootstrap where respondents are resampled with replacement and uncertainty estimates are calculated based on the empirical distribution of the AMCE over the resamples

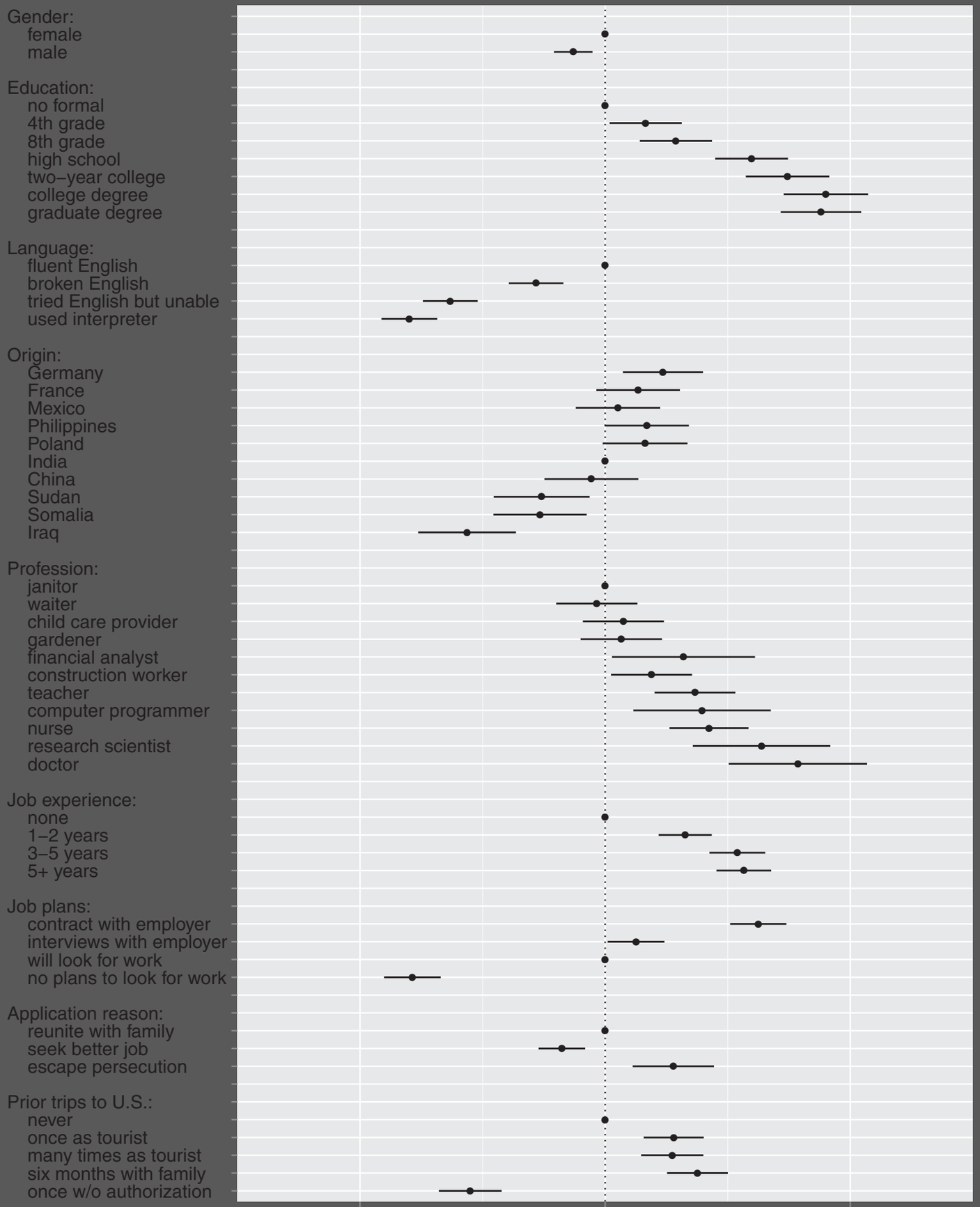
Example: Candidate Experiment

- 3,466 rated profiles – 1,733 pairings
- 311 respondents
- design yields 186,624 possible profiles – far exceeds number of completed tasks
- Respondents rated each candidate profile on a seven-point scale, where 1 indicates that the respondent would "never support" the candidate and 7 indicates that she would "always support" the candidate
- rescaled to 0 and 1

AMCE for candidate age levels

$$\begin{aligned}\text{rating}_{ijk} =& \beta_0 + \beta_1[\text{age}_{ijk} = 75] + \beta_2[\text{age}_{ijk} = 68] + \\ & \beta_3[\text{age}_{ijk} = 60] + \beta_4[\text{age}_{ijk} = 52] + \\ & \beta_5[\text{age}_{ijk} = 45] + \epsilon\end{aligned}$$

- the reference category is 36 years old
- βs are estimators for AMCE for ages 68, 75, etc. compared to 36


Gender:
female
male
Education:
no formal
4th grade
8th grade
high school
two–year college
college degree
graduate degree
Language:
fluent English
broken English
tried English but unable
used interpreter
Origin:
Germany
France
Mexico
Philippines
Poland
India
China
Sudan
Somalia
Iraq
Profession:
janitor
waiter
child care provider
gardener
financial analyst
construction worker
teacher
computer programmer
nurse
research scientist
doctor
Job experience:
none
1–2 years
3–5 years
5+ years
Job plans:
contract with employer
interviews with employer
will look for work
no plans to look for work
Application reason:
reunite with family
seek better job
escape persecution
Prior trips to U.S.:
never
once as tourist
many times as tourist
six months with family
once w/o authorization
–.2
0
.2
Change in Pr(Immigrant Preferred for Admission to U.S.)

CACE External Validity

- Hypethetical versus Behavioral choices
- Treatment versus Control
- Hainmueller et al 2015 PNAS
- Behavioural data from Swiss referendum
- Results matched to conjoint experiment

Swiss Behavioural Data

- municipalities used referendums to vote on the naturalization applications of immigrants.

- voters received a voting leaflet with a short description of the applicant, including information about attributes, such as age, sex, education, origin, language skills, and integration status

- voters then cast a secret ballot to accept orreject individual applicants

- these voting data yield an accurate measure of the revealed preferences of the voters what components of manipulation produce observed effect

Centre for Experimental Social Sciences

Conjoint Experiment Data

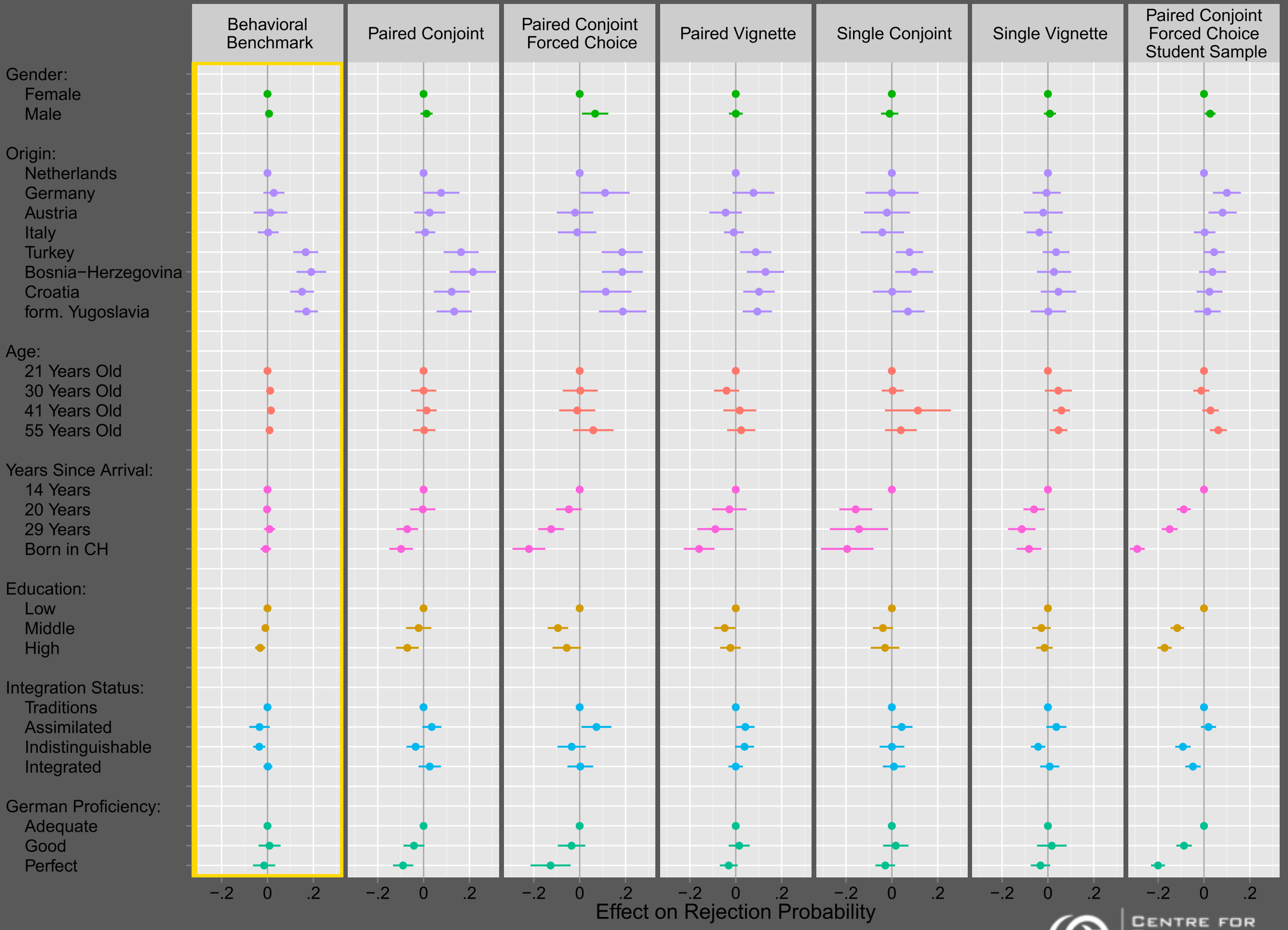
- respondents are presented with profiles of immigrants and then asked to decide on their application for naturalization

- list of attributes matches attributes voters saw no the voting leaflets distributed for the referendums — presented in same order as on the original leaflets

- each respondent is randomly assigned to one of five different designs and asked to complete 10 choice tasks

Experimental Designs

Designs	Profiles
single-profile vignette	accept/reject single profile
paired-profile vignette	accept/reject two profiles
single-profile conjoint	name/value of attributes
paired-profile conjoint	accept/reject 2 applicants
paired-profile conjoint	accept/reject 1 of 2 applicants



Centre for Experimental Social Sciences

Conjoint: Causal Interaction

- Egami et al 2016
- Average Marginal Interaction Effect (AIME)
- Does not depend on choice of baseline
- Non-parametrically estimated using ANOVA regression with weighted zero-sum constraints
- Reduces false discovery rate & facilitates interpretation

AME and AMIE Estimation

$$Y_i(\mathbf{t}) \;=\; \mu + \sum_{j=1}^{4}\sum_{\ell=0}^{L_j-1} \beta_\ell^j \mathbf{1}\{t_{ij}=\ell\} + \sum_{j=1}^{4}\sum_{j'\neq j}\sum_{\ell=0}^{L_j-1}\sum_{m=0}^{L_{j'}-1} \beta_{\ell m}^{jj'} \mathbf{1}\{t_{ij}=\ell, t_{ij'}=m\} + \epsilon_i(\mathbf{t}) \tag{14}$$

$$\begin{aligned}\Pr(Y_i(\mathbf{T}_i^*) > Y_i(\mathbf{T}_i^\dagger) \mid \mathbf{T}_i^*, \mathbf{T}_i^\dagger) \;&=\; \tilde{\mu} + \sum_{j=1}^{4}\sum_{\ell=0}^{L_j-1} \beta_\ell^j (\mathbf{1}\{T_{ij}^*=\ell\} - \mathbf{1}\{T_{ij}^\dagger=\ell\}) \\ &+ \sum_{j=1}^{4}\sum_{j'\neq j}\sum_{\ell=0}^{L_j-1}\sum_{m=0}^{L_{j'}-1} \beta_{\ell m}^{jj'} (\mathbf{1}\{T_{ij}^*=\ell, T_{ij'}^\dagger=m\} - \mathbf{1}\{T_{ij}^*=\ell, T_{ij'}^\dagger=m\})\end{aligned}$$

	Range	Selection prob.
AME		
Record	0.122	1.00
Coethnicity	0.053	1.00
Platform	0.023	0.93
Degree	0.000	0.33
AMIE		
Coethnicity × Record	0.053	1.00
Record × Platform	0.030	0.92
Platform × Coethnic	0.008	0.64
Coethnicity × Degree	0.000	0.62
Platform × Degree	0.000	0.35
Record × Degree	0.000	0.09

Factor	AME	Selection prob.
`Record`		
Yes/Village	0.122	⟩ 0.71
Yes/District	0.122	⟩ 0.77
Yes/MP	0.101	⟩ 1.00
No/Village	0.047	⟩ 0.74
No/District	0.051	⟩ 0.74
No/MP	0.047	⟩ 1.00
No/Businessman	base	
`Platform`		
Jobs	−0.023	⟩ 0.56
Clinic	−0.023	⟩ 0.94
Education	base	
`Coethnicity`	0.053	1.00
`Degree`	0.000	0.33

Example

$$
\begin{aligned}
& \underbrace{\tau(\texttt{Coethnic, No/Business};\texttt{Non-coethnic, No/MP})}_{-2.4} \\
= \; & \underbrace{\psi(\texttt{Coethnic};\texttt{Non-coethnic})}_{5.3} + \underbrace{\psi(\texttt{No/Business};\texttt{No/MP})}_{-4.7} \\
& + \underbrace{\pi(\texttt{Coethnic, No/Business};\texttt{Non-coethnic, No/MP})}_{-3.0}
\end{aligned}
$$

Online Experiments: Modes & Subject Pools

MTurk: Static External Validity

- Berinsky et al 2012
- whether estimated (average) treatment effects are accurate assessments of treatment effects for other samples
- whether these estimates are reliable assessments of treatment effects for the same sample outside the MTurk setting.

Table 1 Task title, compensation, and speed of completion for selected MTurk studies

Task title	*Date launched*	*Number of subjects*	*Pay per subject*	*Mean minutes per subject*	*Completions per day*								
					1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Answer a survey about current affairs and your beliefs	January 5, 2010	490	$0.15	7	116	64	41	40	27	36	15	11	15
2- to 3-Min survey for political science research	March 16, 2010	500	$0.25	2	210	68	37	55	53	64	18		
4-Min survey for political science research	April 26, 2010	500	$0.40	4	298	105	79	18					
3-Min survey for political science research	April 29, 2010	200	$0.25	1	200								
3- to 4-Min survey for political science research	May 17, 2010	150	$0.45	2	150								
7- to 9-Min survey	June 24, 2010	400	$0.75	6	400								
5- to 7-Min survey	June 28, 2010	400	$0.75	5	321	79							
5- to 7-Min survey	July 3, 2010	400	$0.50	3	256	115	29						
2- to 3-Min survey	July 16, 2010	200	$0.25	3	200								

Note. The remaining subjects for the January 5, 2010, study were recruited as follows: Day 10 (11), Day 11 (10), Day 12 (17), Day 13 (22), Day 14 (29), and Day 15 (36).

	Convenience Samples				
				Adult samples (Berinsky and Kinder 2006)	
Demographics	*MTurk*	*Student samples (Kam et al. 2007)*	*Adult sample (Kam et al. 2007)*	*Experiment 1: Ann Arbor, MI*	*Experiment 2: Princeton, NJ*
Female	60.1% (2.1)	56.7% (1.3)	75.7% (4.1)	66.0%	57.1%
Age (mean years)	32.3 (0.5)	20.3 (8.2)	45.5 (.916)	42.5	45.3
Education (mean years)	14.9 (0.1)	—	5.48 (1.29)	15.1	14.9
White	83.5 (1.6)	42.5	82.2 (3.7)	81.4	72.4
Black	4.4 (0.9)			12.9	22.7
Party identification					
Democrat	40.8 (2.1)			46.1	46.5
Independent	34.1 (2.0)			20.6	17.6
Republican	16.9 (1.6)			16.3	25.8
None/other	8.2 (1.2)			17.0	10.1
N	484–551	277–1428	109	141	163

Note. Percentages except for age and education with SEs in parentheses. Adult sample from Kam et al. (2007) is for campus employee participants from their Table 1, Column 1. MTurk survey is from February/March 2010.

Table 3 Comparing MTurk sample demographics to Internet and face-to-face samples

	Internet sample		*Face-to-face samples*	
	MTurk	*ANESP*	*CPS 2008*	*ANES 2008*
Female	60.1% (2.1)	57.6% (0.9)	51.7% (0.2)	55.0% (1.3)
Education (mean years)	14.9 (0.1)	16.2 (0.1)	13.2 (0.0)	13.5 (0.1)
Age (mean years)	32.3 (0.5)	49.7 (0.3)	46.0 (0.1)	46.6 (0.5)
Mean income	$55,332 ($1,659)	$69,043 ($794)	$62,256 ($130)	$62,501 ($1,467)
Median income	$45,000	$67,500	$55,000	$55,000
Race				
White	83.5 (1.6)	83.0 (0.7)	81.2 (0.1)	79.1 (0.9)
Black	4.4 (0.9)	8.9 (0.5)	11.8 (0.1)	12.0 (0.6)
Hispanic	6.7 (1.1)	5.0 (0.4)	13.7 (0.1)	9.1 (0.5)
Marital status				
Married	39.0 (2.1)	56.8 (0.9)	55.7 (0.2)	50.1 (1.3)
Divorced	7.1 (1.1)	12.1 (0.6)	10.2 (0.1)	12.9 (0.8)
Separated	2.5 (0.7)	1.3 (0.2)	2.1 (0.1)	2.9 (0.4)
Never married	50.6 (2.1)	14.2 (0.6)	25.7 (0.2)	26.2 (1.1)
Widowed	0.7 (0.4)	4.9 (0.4)	6.3 (0.1)	7.8 (0.6)
Housing status				
Rent	52.7 (2.3)	14.3(0.1)		32 (1.2)
Own home	47.3 (2.3)	80.8 (0.8)		66.1 (1.2)
Religion				
None	41.8 (2.1)	13.1 (0.8)		26.9 (1.2)
Protestant	20.7 (1.7)	38.7 (1.4)		28.2 (1.2)
Catholic	16.5 (1.6)	22.9 (1.0)		17.5 (1.0)
Jewish	4.4 (0.9)	3.0 (0.4)		1.2 (0.3)
Other	16.5 (1.6)	22.2 (1.0)		26.2 (1.1)

Table 4 Comparing MTurk sample political and psychological measures to Internet and face-to-face samples

	Internet sample		*Face-to-face samples*	
	MTurk	*ANESP*	*CPS 2008*	*ANES 2008*
Registration and turnout				
Registered	78.8% (1.7)	92.0% (0.7)	71.0% (0.2)	78.2% (1.1)
Voter turnout 2008	70.6 (2.0)	89.8 (0.5)	63.6 (0.2)	70.4 (1.1)
Party identification (mean on 7-point scale, 7 = Strong Republican)	3.48 (0.09)	3.90 (0.05)		3.70 (0.05)
Ideology (mean on 7-point scale, 7 = Strong conservative)	3.39 (0.09)	4.30 (0.05)		4.24 (0.04)
Political Interest (mean on 5-point scale, 5 = Extremely interested)	2.43 (0.04)	2.71 (0.02)		2.93 (0.03)
Political knowledge (% correct)				
Presidential succession after Vice President	70.0 (1.3)	65.2 (2.0)		
House vote percentage needed to override a veto	81.3 (1.7)	73.6 (1.3)		
Number of terms to which an individual can be elected president	96.2 (0.8)	92.8 (0.7)		
Length of a U.S. Senate term	45.0 (2.1)	37.5 (1.3)		
Number of Senators per state	85.4 (1.5)	73.2 (1.2)		
Length of a U.S. House term	50.1 (2.1)	38.9 (1.3)		
Average	71.3	63.5		
Need for cognition (mean on 0–1 scale)	.625 (0.012)	.607 (0.006)		.559 (0.009)
Need to evaluate (mean on 0–1 scale)	.628 (0.008)	.579 (0.004)		.558 (0.005)
N	506–699	1,466–2,984	92,360	1,058–2,323

Note. Means with SEs in parentheses. CPS 2008 and ANES 2008 are weighted. Political measures are from the February/March 2010 MTurk survey ($N = 551$). Need for Cognition and Need to Evaluate are from the May 2011 MTurk survey ($N = 699$). Tests of statistical significance of differences across samples appear in the Supplementary data.

Table 6 Replication of Table 2 by Kam and Simas (2010)—risk acceptance and preference for the probabilistic outcome

	Kam and Simas (2010)			*MTurk replication*		
	(H1a) Mortality frame and risk acceptance	*(H1b) Adding controls*	*(H2) Frame × Risk acceptance*	*(H1a) Mortality frame and risk acceptance*	*(H1b) Adding controls*	*(H2) Frame × Risk acceptance*
Mortality frame in Trial 1	1.068 (0.10)	1.082 (0.10)	1.058 (0.29)	1.180 (0.10)	1.180 (0.10)	1.410 (0.31)
Risk acceptance	0.521 (0.31)	0.628 (0.32)	0.507 (0.48)	0.760 (0.29)	0.780 (0.31)	0.990 (0.42)
Female		0.105 (0.10)			−0.018 (0.11)	
Age		0.262 (0.22)			0.110 (0.31)	
Education		−0.214 (0.20)			0.025 (0.23)	
Income		0.205 (0.23)			−0.024 (0.23)	
Partisan ideology		0.038 (0.19)			0.006 (0.15)	
Risk acceptance × Mortality frame			0.023 (0.62)			−0.450 (0.58)
Intercept	−0.706 (0.155)	−0.933 (0.259)	−0.700 (0.227)	−1.060 (−0.170)	−1.100 (−0.290)	−1.190 (−0.230)
lnL	−453.185	−450.481	−453.184	−409.740	−409.662	−409.439
$p > \chi^2$	0.000	0.000	0.000	0.000	0.000	0.000
N	752	750	752	699	699	699

Note. Entries are probit coefficients with SEs in parentheses. Dependent variable is Preference for the Probabilistic Outcome (0 = deterministic outcome; 1 = probabilistic outcome). All independent variables are scaled to range from 0 to 1. MTurk survey is from May 2010. None of the differences between coefficients across studies are statistically significant (see the Supplementary data).

MTurk: Internal Validity

- the possibility that subjects violate treatment assignment by participating in a given task more than once.
- subject inattentiveness, in which case some subsets of the sample do not attend to the experimental stimuli and are effectively not treated

MTurk: Internal Validity Evidence

- Participating in experiments multiple times
- 7 out of 551 IP addresses (2.4%) produced 2 responses
- Attention, demand and subject motivation
- questions to confirm attentiveness – 60% recalled political office of person described in vignette story
- 95% prior approval rating

Conclusions

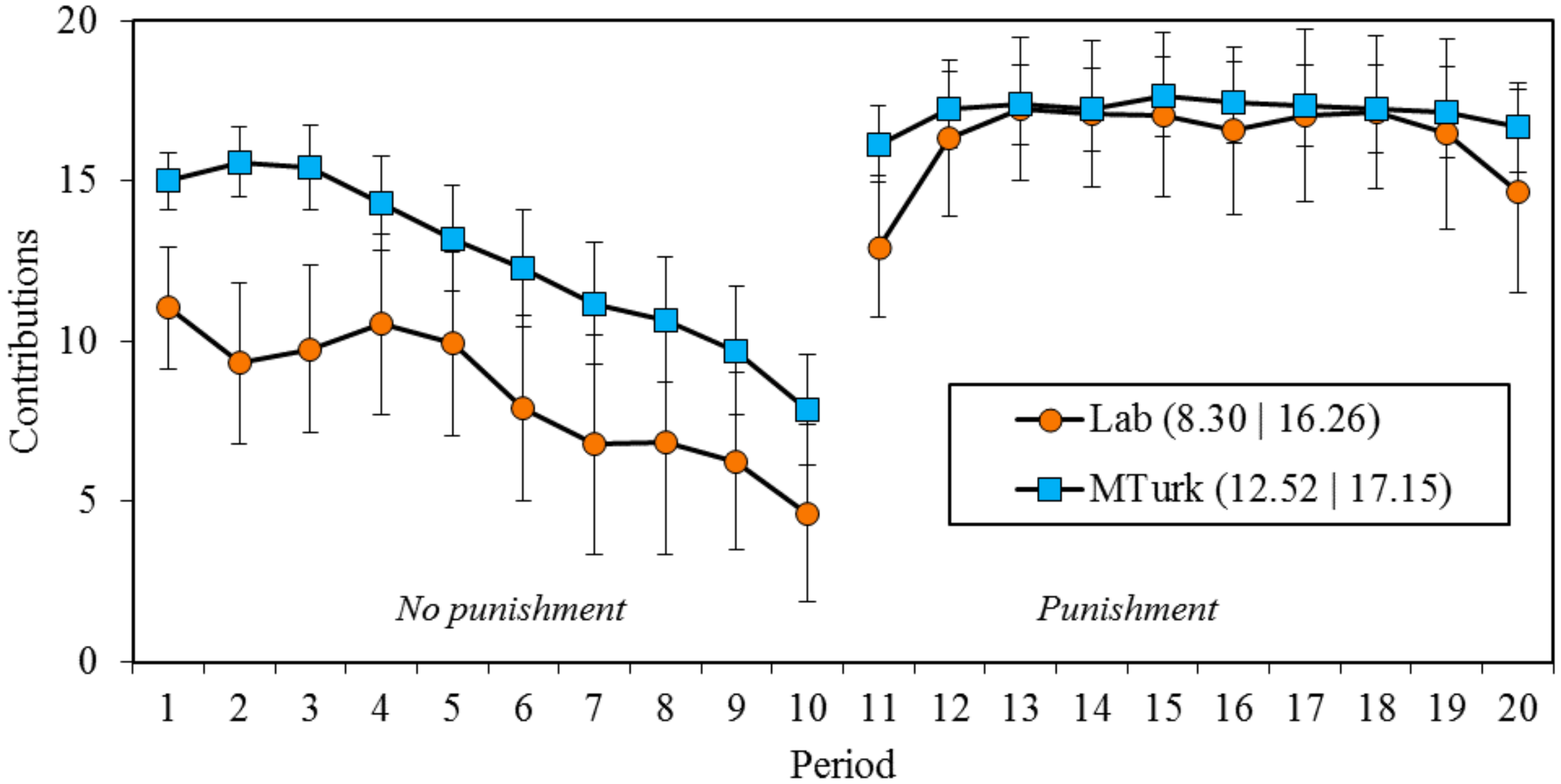
- the demographic characteristics of domestic MTurk users are more representative and diverse than the corresponding student and convenience samples typically used in experimental studies

- they replicate experimental studies previously conducted using convenience and nationally representative samples, finding that the estimates of average treatment effects are similar in the MTurk and original samples.

- they find that potential limitations to using MTurk to recruit subjects and conduct research – in particular, concerns about heterogeneous treatment effects, subject attentiveness, and the prevalence of habitual survey taker – are not large problems in practice.

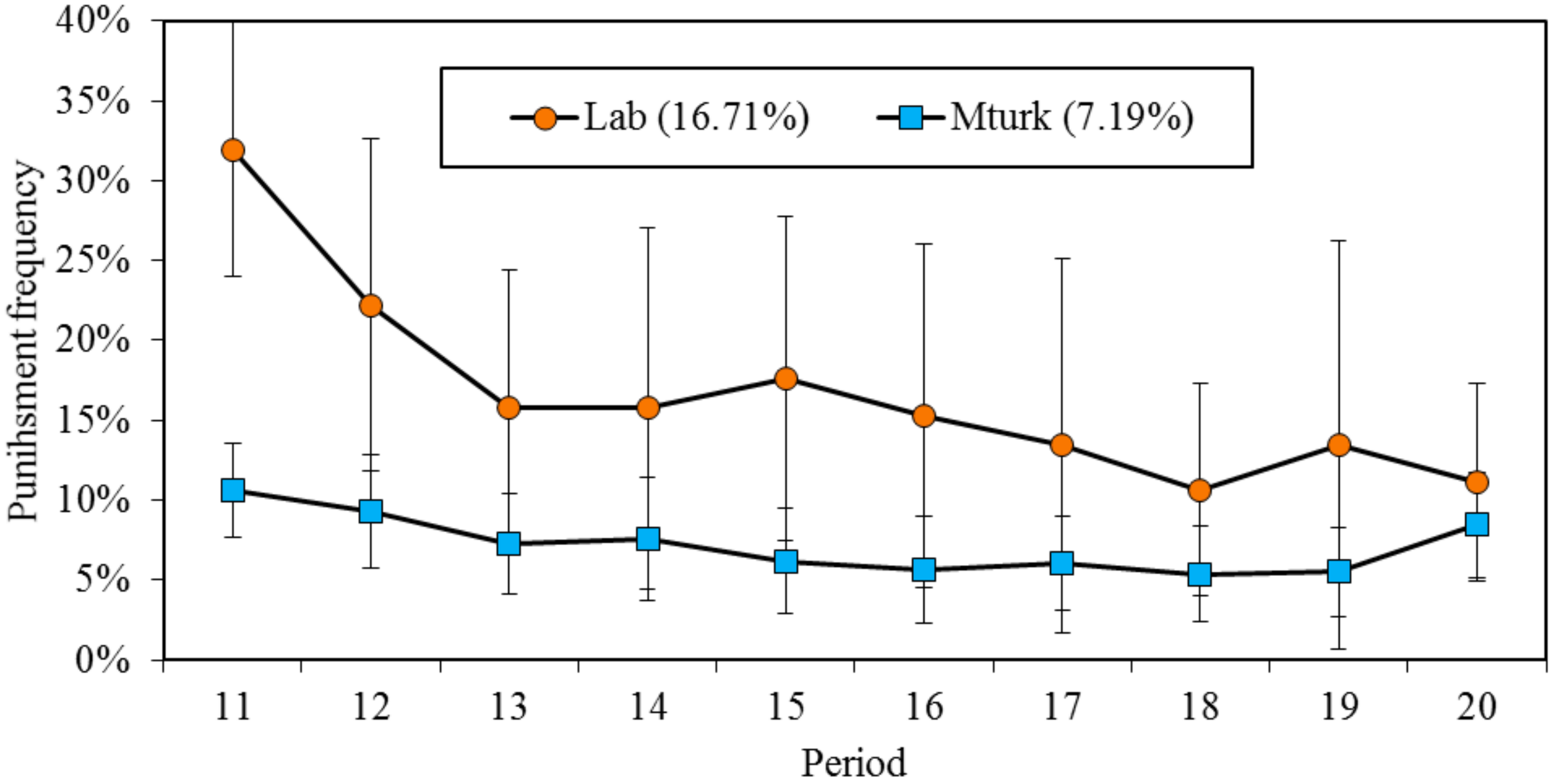
MTurk: Dynamic External Validity

- Arechar et al 2017
- Lab Experiment versus MTurk ``Hot'' Experiment
- Public Goods Game



	Contributions to the public good					
	No punishment			Punishment		
	Laboratory	MTurk	Pooled	Laboratory	MTurk	Pooled
Period	-0.900***	-1.074***	-1.037***	1.139	0.514*	0.682**
	(0.309)	(0.187)	(0.160)	(0.710)	(0.289)	(0.282)
Final period	-3.400	-2.292**	-2.512***	-10.203**	-4.184**	-5.795***
	(2.253)	(0.958)	(0.881)	(4.881)	(1.688)	(1.797)
MTurk			5.421***			4.193
			(1.867)			(4.904)
Constant	10.470***	17.046***	11.402***	25.980***	35.272***	29.601***
	(1.592)	(0.624)	(1.650)	(3.898)	(3.792)	(4.232)
N	720	2480	3200	720	2480	3200
F	8.75	33.66	34.45	2.19	3.12	3.75

Table 2 Cooperation dynamics. Tobit estimation with left-censoring for 'No punishment' and right-censoring for 'Punishment'. 'Period' is period number; 'Final period' is a dummy for last period; 'MTurk' is a dummy for the MTurk sample. Robust standard errors clustered on groups; * $p<0.1$, ** $p<0.05$, *** $p<0.01$.



	Participant's drop out in period t (0=no; 1=yes)				
	Pooled data			*Without punishment*	*With punishment*
	(1)	(2)	(3)	(4)	(5)
Punishment available	0.056	0.362	0.107		
	(0.598)	(0.612)	(0.611)		
Period	-0.093*	-0.118**	-0.094*	-0.265***	-0.150*
	(0.051)	(0.053)	(0.053)	(0.080)	(0.082)
First period	2.484***	2.375***	2.554***		
	(0.377)	(0.376)	(0.382)		
Earnings		-0.002	0.011		
		(0.143)	(0.143)		
Group member(s) dropped out in previous period			1.890***	3.636***	2.034***
			(0.382)	(0.394)	(0.573)
Relative average contribution				0.010	-0.082
				(0.033)	(0.053)
Relative average punishment received					-0.104
					(0.092)
Relative average punishment given					-0.204
					(0.214)
Constant	-4.064***	-3.979***	-4.220***	-3.519***	-4.282***
	(0.317)	(0.318)	(0.328)	(0.466)	(0.472)
N	8334	8327	8332	3539	3527
AIC	893.56	877.20	860.27	325.98	302.30

Dynamic Multi-modes

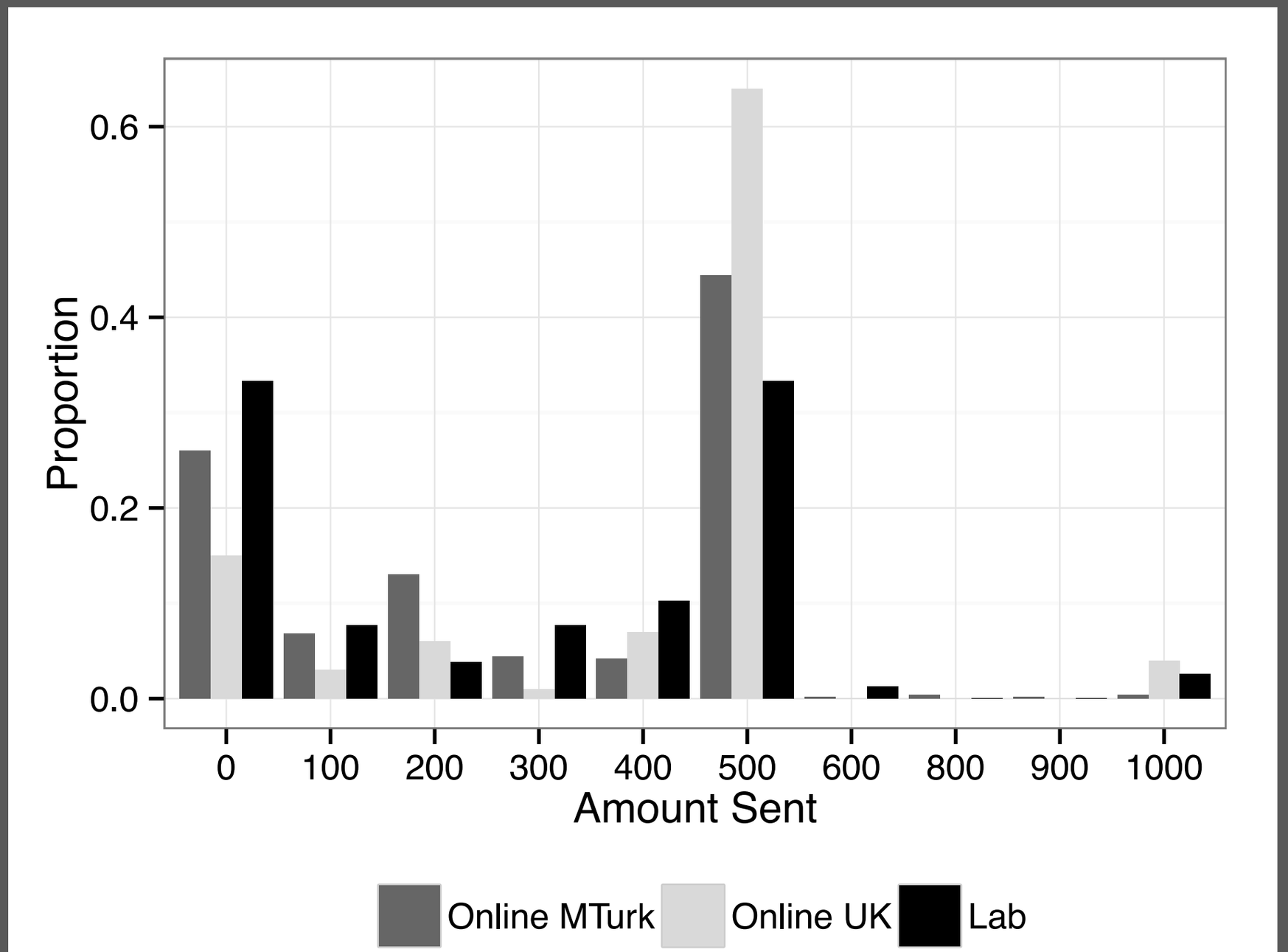
- Duch et al 2017
- Modes: Online versus Lab Experiment
- Subject Pools:
- Public Goods Game + RET

Table 1: Summary of Mode and Subject Pool Effects

Mode	Lab Subject Pool	Online Subject Pool	Subject Pool Characteristics
Lab BD	Lab-BD	NA	NA
Online Non-synchronic	Online Lab-BD	CESS Online-BD	Estimated
Online Non-synchronic (M-Turk)	NA	MTurk-BD	NA
Lab DS	Lab-DS	NA	NA
Online Synchronic (M-Turk)	NA	MTurk-DS	NA
Online Synchronic	Online Lab-DS	NA	Estimated
Mode Effect	Estimated	Estimated	

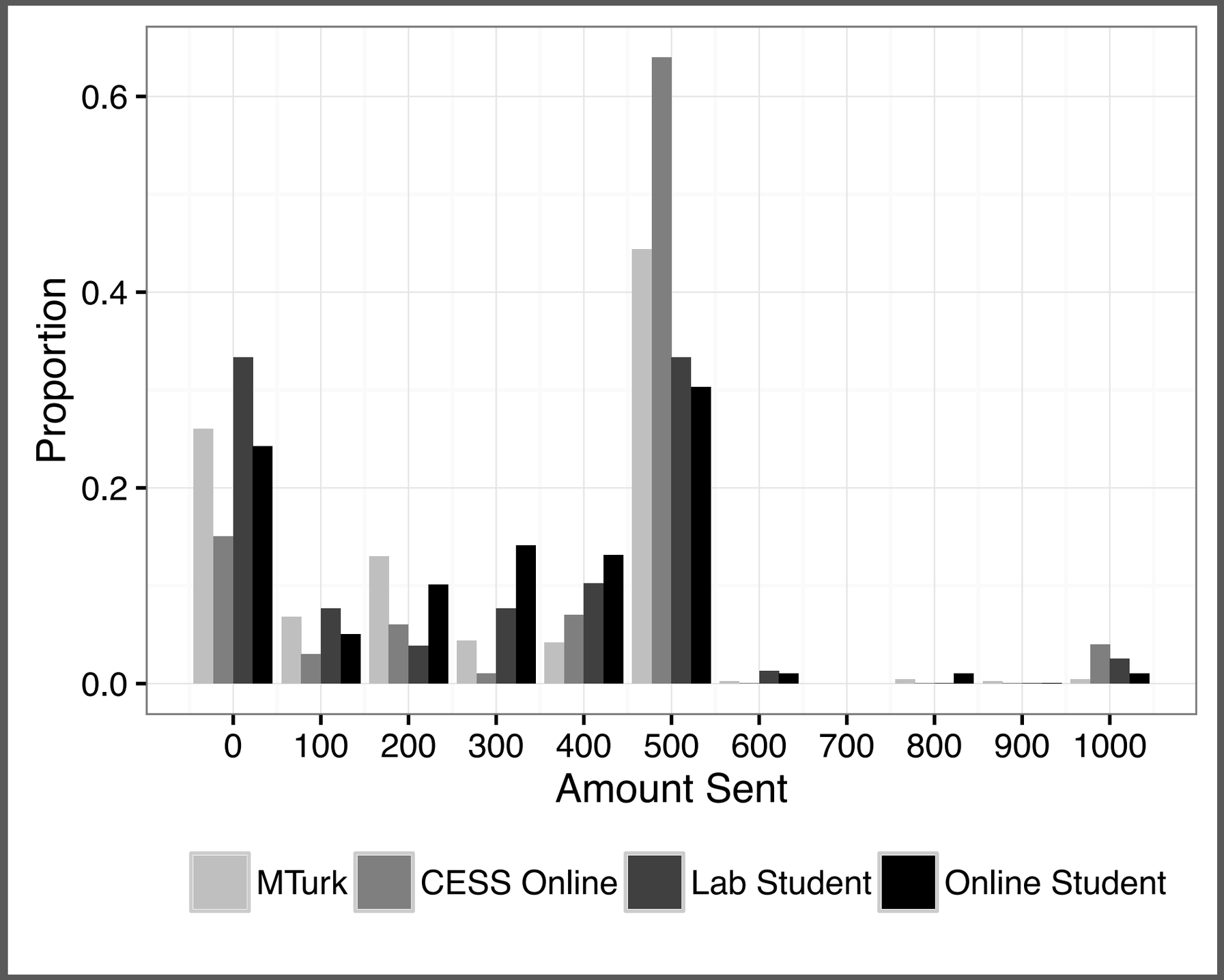
Note Mode and Subject Pool Effects.

Duch & Solaz

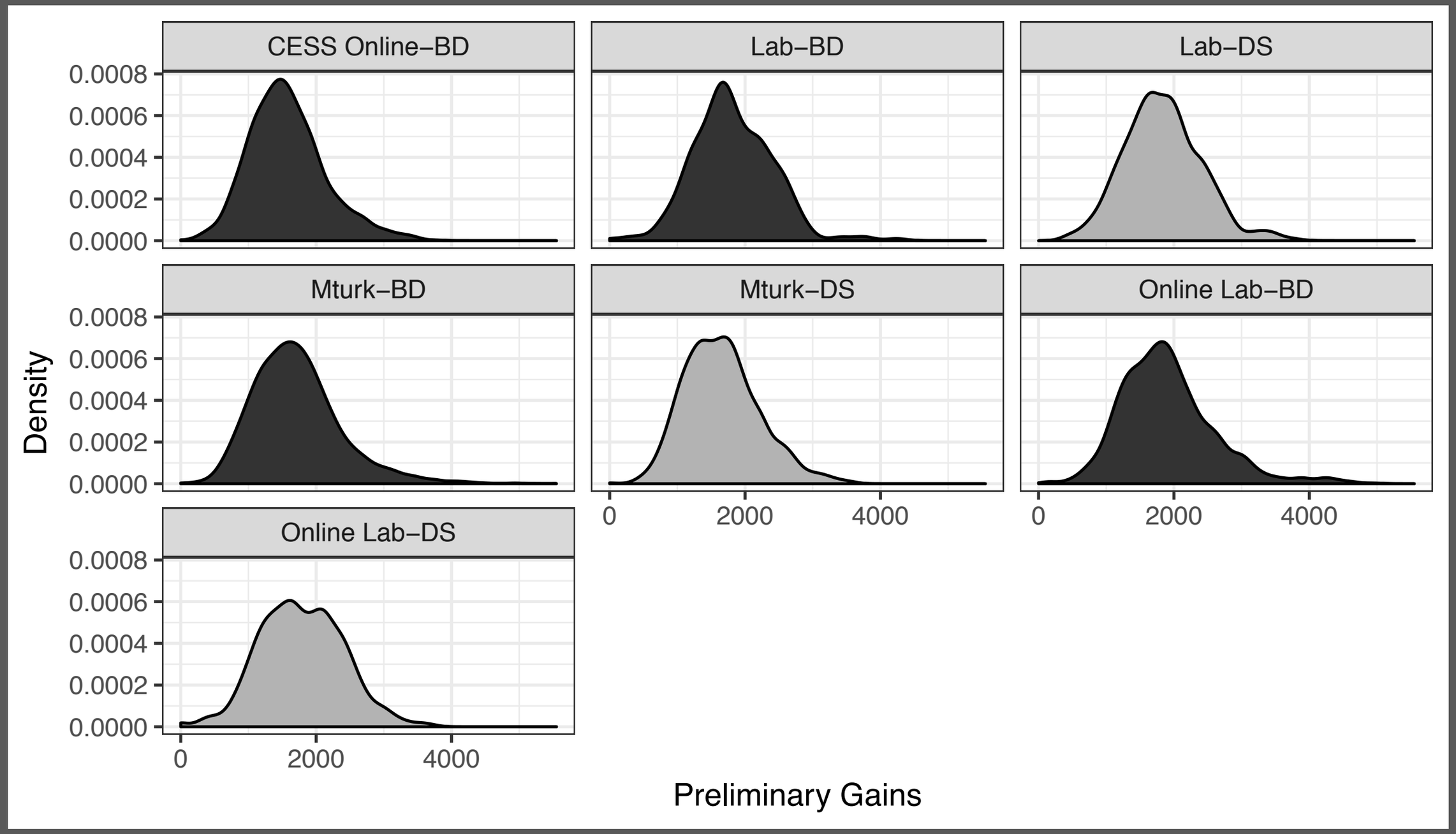


Centre for Experimental Social Sciences

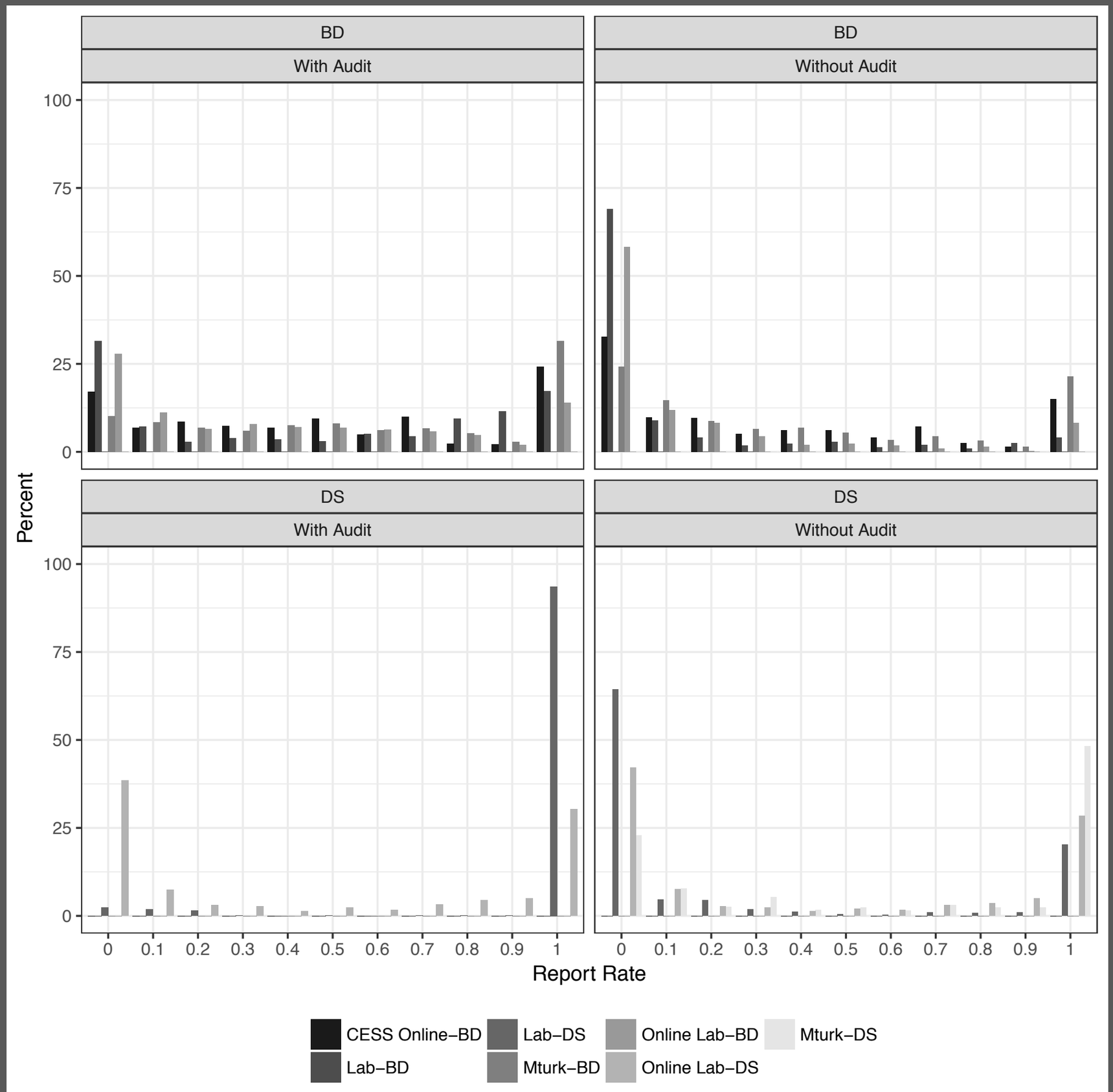
Duch & Beramindi



Real Effort Task



Report Rate



CENTRE FOR EXPERIMENTAL SOCIAL SCIENCES

Multivariate

	CESS Online-BD	Mturk BD	Lab BD	Online Lab-BD	Mturk DS	Lab DS	Online Lab-DS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# of Additions	−0.001	−0.006***	−0.023***	−0.016***	−0.013***	−0.015***	−0.011***
	(0.006)	(0.002)	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)
Middle Tax Bracket	0.046	0.033*	−0.069*	0.009		0.028	
	(0.043)	(0.017)	(0.036)	(0.030)		(0.021)	
Low Tax Bracket	0.083	0.062***	−0.054	0.073*	−0.062***	0.088***	0.130***
	(0.057)	(0.021)	(0.047)	(0.037)	(0.022)	(0.021)	(0.018)
No Audit	−0.184***	−0.178***	−0.318***	−0.205***		−0.705***	−0.038**
	(0.027)	(0.011)	(0.024)	(0.020)		(0.016)	(0.018)
Dictator Game Giving	0.318***	0.311***	0.271***	0.158***	0.429***	0.323***	0.452***
	(0.066)	(0.026)	(0.050)	(0.051)	(0.044)	(0.043)	(0.040)
Integrity Score	−0.116	−0.366***	−0.220*	−0.001	−0.010	0.115*	−0.176***
	(0.113)	(0.044)	(0.117)	(0.086)	(0.066)	(0.059)	(0.059)
Risk Preference	−0.054	−0.048*	−0.200***	−0.196***	0.107**	0.087*	−0.228***
	(0.055)	(0.026)	(0.070)	(0.062)	(0.053)	(0.045)	(0.047)
Constant	0.402***	0.637***	0.916***	0.609***	0.601***	0.944***	0.508***
	(0.113)	(0.039)	(0.115)	(0.084)	(0.045)	(0.051)	(0.042)
Observations	725	4,460	728	968	1,632	1,440	2,249
R^2	0.116	0.135	0.287	0.218	0.075	0.577	0.113
Adjusted R^2	0.107	0.134	0.281	0.212	0.072	0.575	0.111